

半導体漫遊記

325

湯之上隆

2022年11月30日

OpenAIがChatGPTを公開して以降、生成AIの世界のブームはとどまるどころを知らず、ハイテク企業が続々と生成AI開発に乗り出している。その生成AIには、米NVIDIAが設計し、ファウンドリのTSMCが製造している画像処理プロセッサ(GPU)がAI半導体として使われている。その代表的なGPUの「A100」と「H100」は、恐ろしく巨大でとんでもなく高価なチップである。

トランジスタ数が540億個で、1個1万ドルである。またTSMCの4nmで製造される「H100」は、チップサイズが814平方

メートルという。またOpenAIが23年7月に商標登録を行った「GPT-5」には、3.5万個の「H100」が必要と推測されている。

おいてはNVIDIAの「H100」を200万個製造が設計し、TSMCが「H100」の場合には、4nmプロセスで前工程を行い、さらに後工程もTSMCがNVIDIAのGPU用に開発したCoWoS(Chip on Wafer on Substrate)でパッケージングする。

GPUの「H100」に割り当ててもらう必要がある。12.5万枚である。というこは、毎月1万枚強のウエハをNVIDIAのGPUの「H100」に割り当ててもらう必要がある。という問題がと深刻である。そもそもTSMCは前工程が専門のファウンドリであり、後工程の能力がほとんどない。そこに降って湧いたNVIDIAのGPU騒動である。これに對処するため、TSMCは台湾に先進パッケージング工場「AP6」を23年6月に(慌てて)開設した。24年末には現在の2倍の月産2.8万枚の生産能力に達すると報道されているが、そこにたどり着くまでは後工程のキャパシティ不足に悩まされることになるだろう。(微細加工研究所・所長)

NVIDIA、高価なチップ

TSMCの生産キャパ不足

方mm、トランジスタ数が800億個で、1個4万ドルもする。

そして生成AIの開発を満たすことができず、24年には200万個以上出荷しなければならぬという。しかも、それは簡単なことではない。

まず4nmの前工程が問題である。巨大な「H100」は12インチウエハから最大65個しか取得できない。そしてチップサイズと歩留まりは反比例するた

め、歩留まりは25%程度と思われる。となると「H100」は1枚に次世代の「H200」の製造も始まる。要するに、TSMCがどれだけのキャパシテ

TSMCの7nmプロセスで製造される「A100」は、チップサイズが露光限界に近い826平方mm、

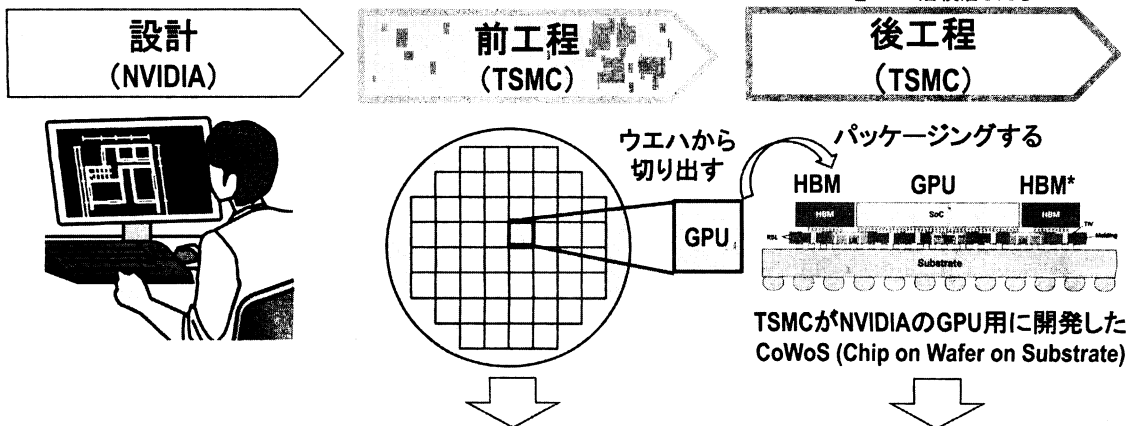
個、Metaが2.1万個、Teslaが7000個、それぞれ「A100」を保有し

つくれる。GPUに

このような「H100

要するに、TSMCが

HBM*→High Bandwidth Memory
DRAMを8~12層積層したもの



チップサイズが巨大(814~826平方mm)
N7~N5で歩留まり悪く、チップ価格高い
「A100」が1万ドル、「H100」が4万ドル

TSMCのキャパシティが足りない
6月に3D IC専用Fabを開設

NVIDIAのGPU製造に関するボトルネック

もTSMCは前工程が専門のファウンドリであり、後工程の能力がほとんどない。そこに降って湧いたNVIDIAのGPU騒動である。これに對処するため、TSMCは台湾に先進パッケージング工場「AP6」を23年6月に(慌てて)開設した。24年末には現在の2倍の月産2.8万枚の生産能力に達すると報道されているが、そこにたどり着くまでは後工程のキャパシティ不足に悩まされることになるだろう。(微細加工研究所・所長)